
Delaware Webinar

Evaluation 101

Energy Efficiency Program Evaluation

By
Nick Hall
TecMarket Works

February 8, 2012

Workshop Objectives

Create a greater understanding of evaluation, evaluation issues and the evaluation process

Address Delaware evaluation-related questions, issues, concerns, needs.



What we will cover in 3 hours

1. History of evaluation
2. Key definitions
3. Evaluation Framework (why needed)
4. Evaluation as portfolio management tool
5. General what, when and why of evaluation
6. Impact evaluation
7. What is EM&V
8. Net to Gross and Attribution
9. Process Evaluation
10. Market Effects Evaluation
11. Cost effectiveness
12. Evaluation plans and planning

History of Energy Program Evaluation

- USDOE formed in 1970s – implementing wide range of information programs
- Early evaluation was 100% ex ante and conducted by the program administrators
- These results were very unreliable – not field based, didn't capture the actual results
- Solution was to create EE program evaluation building upon the broader (non energy) evaluation field, applying those same evaluation definitions and standards to EE program evaluation.
- Overtime the approaches have improved to specifically address the unique issues associated with EE/RE/DR/ME
- State approaches have evolved independently with the introduction of utility programs – creating a need for Frameworks and protocols

Key Definitions

- Ex ante: projected (pre-program estimated) savings to be achieved
- Ex post: measured (evaluated) savings achieved
- EM&V: evaluation, measurement and verification.
- Framework: Evaluation policy and operational systems/structures and definitions
- Protocol: prescribed ways of conducting evaluation efforts
- Gross savings: unadjusted savings achieved by all program participant for a program-covered intervention
- Verified gross savings: savings achieved by all program participant for a program-covered intervention adjusted to account for verified instillations

Key Definitions

- Net savings: total savings achieved as a result of a program or portfolio effort
- Freeriders: participants who would have taken the same action at the same time without the program intervention
- Freedivers/spillover: Non-participants that took actions as a result of the program's interventions but who did not participate in any of the programs offerings
- Participant spillover: Participants who repeat the same actions but did not receive another incentive or program service.

Evaluation Topics of Interest

- Evaluation and its role in understanding the adequacy of a portfolio
 - Typically covered at the program level
 - Not typically covered at portfolio level
 - The New York Approach

Evaluation Topics of Interest

- Why is a Framework Needed...
 - Guides all evaluation efforts with regard to:
 - Who, When, Why, How, Under what conditions
 - Covers most everything related to evaluation
 - Makes sure everyone playing by the same rule book
 - Avoids the oops factor:

Framework Topics

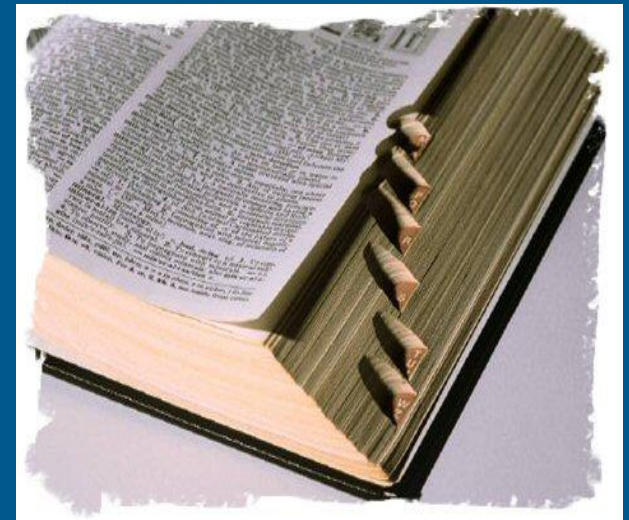
- Framework can cover
 - Approaches to use
 - Objectives and metrics on which to focus
 - Ethics, standards and principles
 - Planning and approval processes
 - Content roles & schedules for TRMs
 - Policies (baselines, net, gross, IPMVP, sampling, timing)
 - Data security and management
 - Customer contact and data collection
 - Planning and budgeting
 - Reporting and report contents
 - Cost effectiveness approach

General What, When, Why of Evaluation



What is Evaluation?

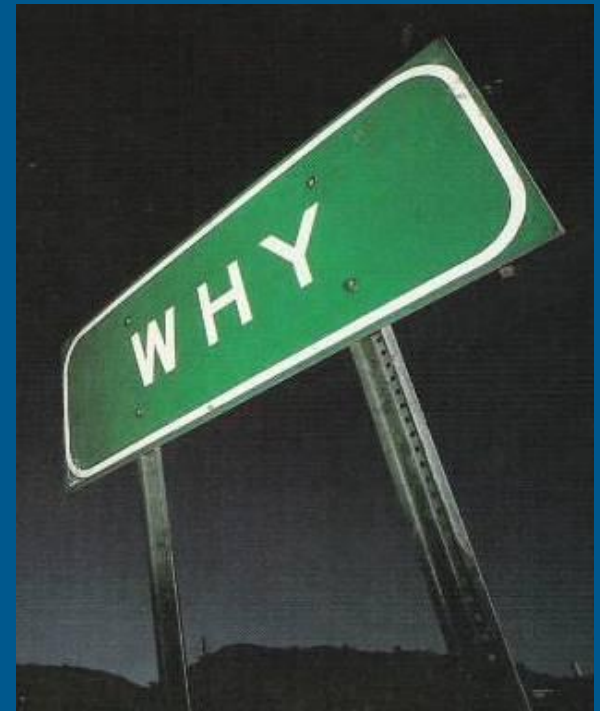
- Evaluation is an **objective systematic process** for assessing an organization's activities in order to quantify the effectiveness, efficiency or effects of those activities for the purpose of documenting performance or making improvements.



Why Evaluate?

Evaluation results can benefit stakeholders by ensuring better and more cost-effective programs!

- Ensure that the program is delivering the benefits that it was designed to produce
- Unbiased independent assessment that supports regulatory process – including cost recovery, administrator compensation, etc.
- Optimize energy and non-energy benefits
- Provide valuable information about program operations



Evaluation Types

- Process evaluation (documents and improves)
- Impact evaluation (short term impacts)
- Market effects (longer term impacts)

Process + Impact + Market effects = a *well rounded* evaluation



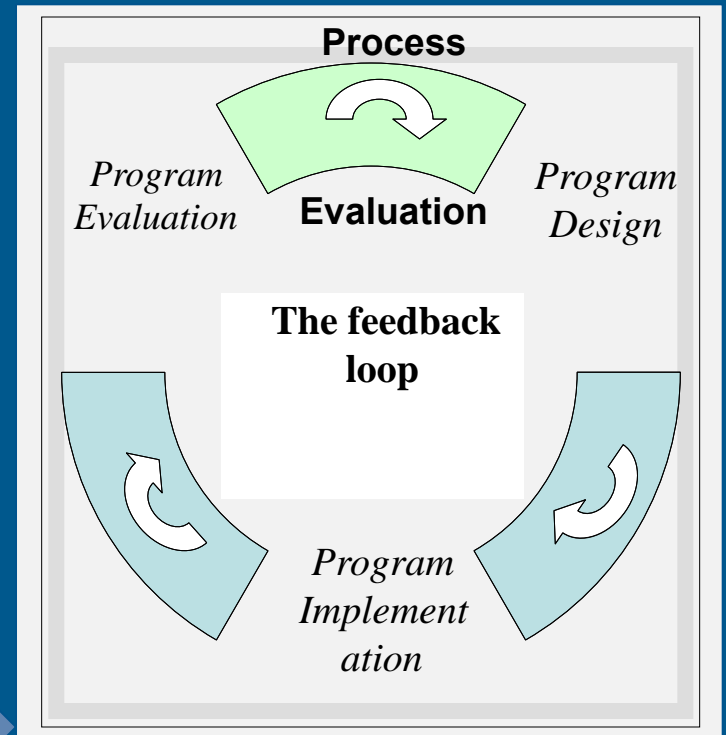
What are we measuring?

- Energy savings
- Demand Savings
- Environmental impacts
- Economic impacts
- Customer satisfaction
- Non-energy benefits
- Technology penetration
- Other program specific research issues



When to Evaluate

- Early enough to be of use!
 - Evaluation create a feedback loop that informs:
 - Program design
 - Program implementation.
- But not too early!
 - Process evaluation (after 6 months)
 - Impact evaluation
 - When there is something to structure into a plan
 - When pre-data is needed
- Regularly within systematic process!
 - The cycle is continuous
 - When a need is identified



Data Collection Primary Methods

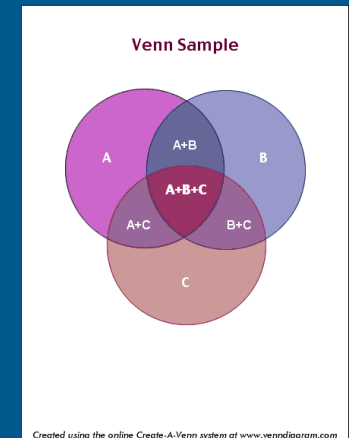
- Surveys (Phone, Mail, Internet, email)
- Focus Groups
- Observation Visits
- Mystery Shopping
- In-depth Interviews
- Site Inspections
- Metering

Important note:
Where possible,
respondents'
data should always
remain
confidential.

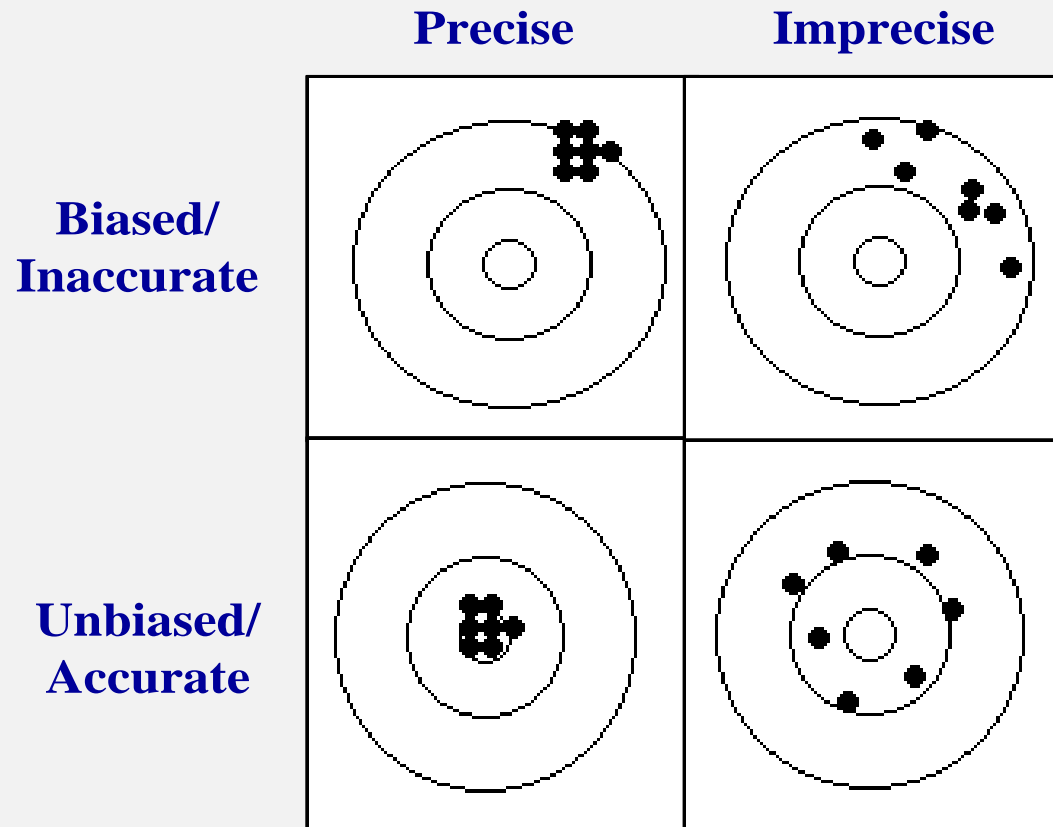
Respondents
should be assured
of confidentiality
before the data
collection.

Sample Design

- Strategy varies by research question and study objectives
- When designing a sampling plan, consider:
 - Population size and distribution
 - Presence of the characteristic being measured and conditions affecting that characteristic
 - Confidence level
 - Precision level
 - Coefficient of variation
 - Effect size



Precision and Bias



Impact Evaluation, Measurement and Verification

The Evaluation Challenge

Evaluation attempts to measure what did not happen.

Measuring invisible energy!

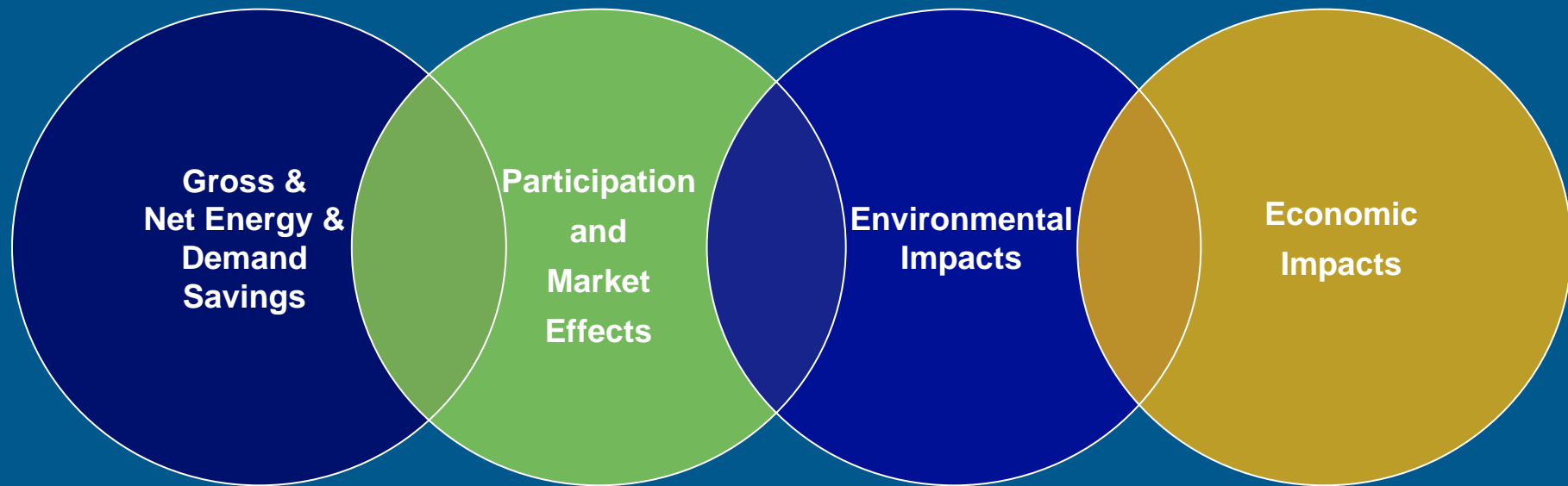
Savings: The difference between energy use after the program and what the energy use would have been without the program

- Not an easy question to answer; we need a
- baseline...



- Nutshell: $\text{Impact} = \text{Actual post} - \text{Actual pre} \pm \text{Adjustment}$

What Do You Measure?

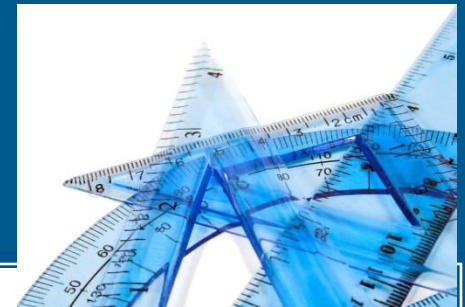


How Do You Measure Impacts?

- Engineering calculations/algorithms
- Billing analysis (utility meter)
- Metered data analysis (evaluation meter)
- Load shape analysis
- Building energy simulation modeling
 - DOE-II

Engineering Approaches

- Engineering calculations use formulas or algorithms to estimate the energy use of equipment before and after installation.
- These approaches are good for projects that do not have a variance in equipment use patterns
- There are many on-line calculators that can be used including EIA DOE, Energy Star, and other web sites
 - Lighting equipment replacements
 - Prescriptive measures such as high efficiency packaged air conditioning
 - Computer and plug-load savings



Billing Analysis

- Uses customer or facility billing data
- May be simple pre-post comparison
 - If pre-measure installation data are not available, such as for new homes, a comparison group is needed.
- May be complex statistical billing analysis
 - Including engineering estimates for installed measures can improve the statistical billing estimates (Statistically Adjusted Engineering (SAE))



Billing Analysis Use



- Use billing analysis when:
 - There is a sufficient number of sample points
 - Sufficient historical data
 - The expected energy savings is “5” percent or more of electric bill
 - There is good data on dates measures were installed and information on specific measures
 - Billing data is relatively clean

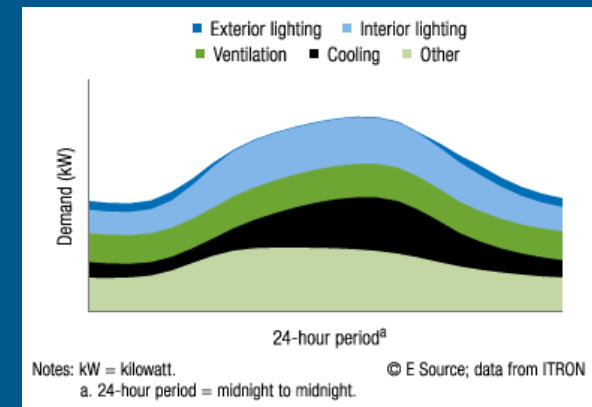
Metered Data Analysis

- Metering end use loads can be the most direct and accurate method for measuring changes in energy consumption
 - Used selectively due to the cost of the equipment and the labor to install/remove the meters
 - Not easily transferred from other climates and service areas



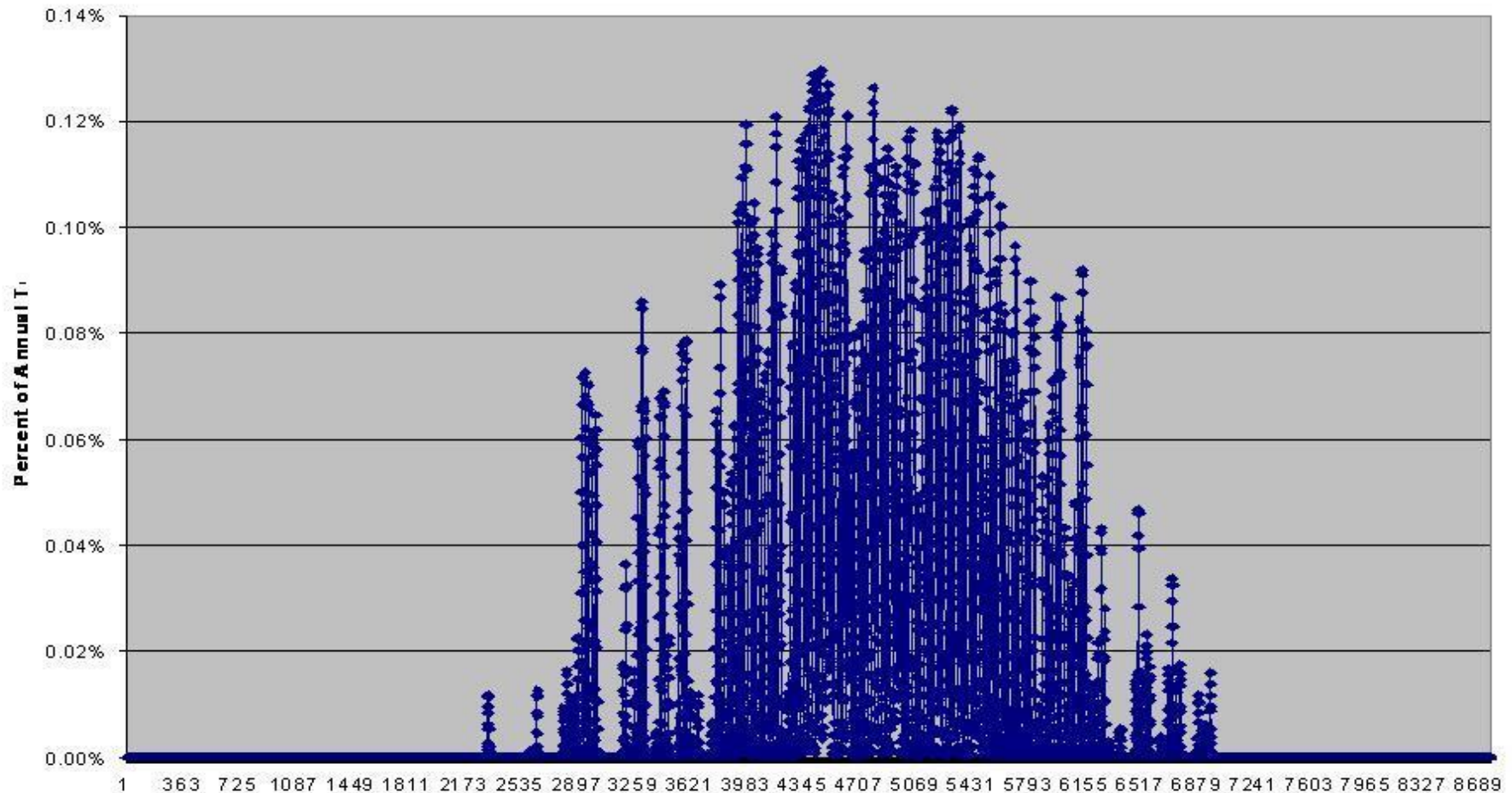
Load Shape Analysis

- Load shape analysis may rely on secondary, as well as primary metered data, to develop end-use load shapes to estimate peak demand or energy savings
- Critical to the evaluation of programs designed to reduce demand or shift loads (demand response programs)



8760 hours (annual) load example

Air conditioning load



Building Simulation Tools

- Engineering estimates of savings may also be based on building simulation models or energy analysis software such as DOE-2, FEDS, EZ Sim, PowerDOE, eQuest, Trace 700, TREAT, and many others
- Billing data useful to calibrate models for specific buildings/typical homes, etc.

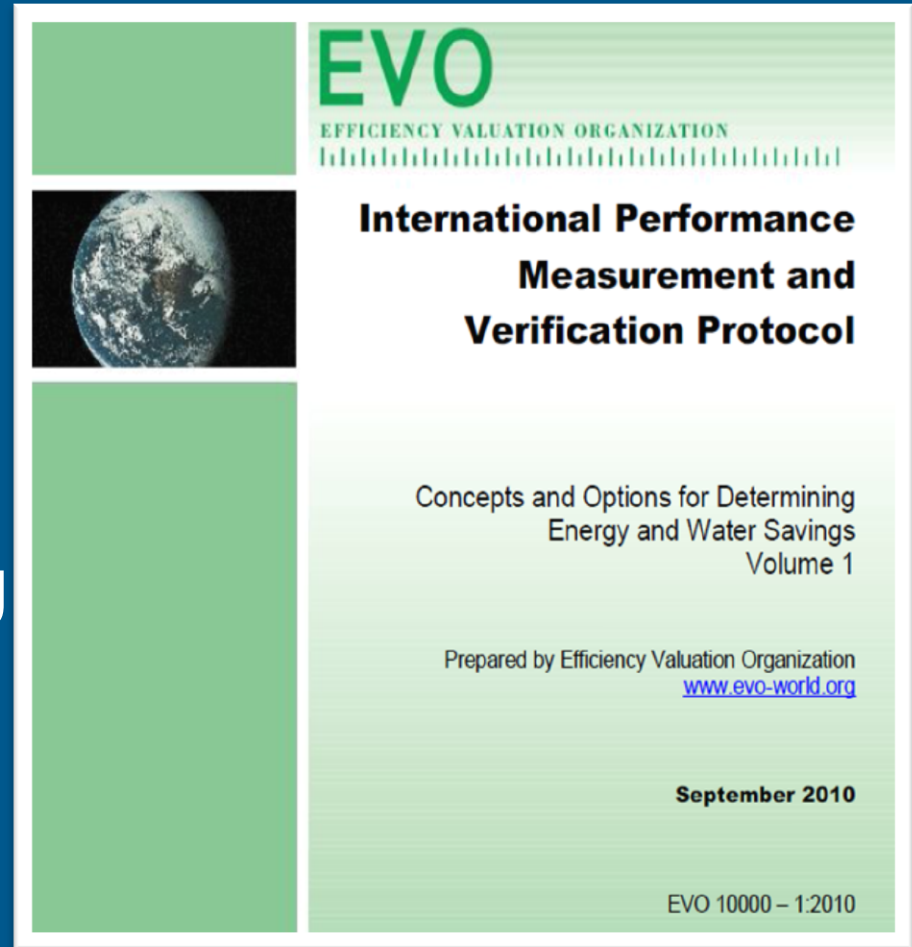


So....How does Impact **E**valuation Differ from **M**easurement and **V**erification?



The M is Guided by the IPMVP

- *International Performance Measurement and Verification Protocol (IPMVP)*
- Framework to determine energy and water savings resulting from the implementation of an energy efficiency project



Definition of M&V

- **Measurement and Verification (M&V)**
- The **M** is for measurement: the process of collecting on-site measured or observed data from an individual project or group of projects. The data is used to feed an energy impact evaluation. (a function of physics or engineering)
- The **V** is for Verification: the process of verifying that measures are installed, configured and used in a way that reflects the assumptions of the program. The data is used to feed an energy impact evaluation. (a confirmation function)
- The **E** in EM&V is for Evaluation, the analysis conducted to determine the amount of energy impacts. This analysis uses the results of the M&V. (an evaluation function)

Program Attribution (or Net to Gross)

- Program attribution refers to energy impacts that can confidently be attributed to program efforts
- A net-to-gross factor that reflects program attribution is applied to gross program energy savings to get net energy savings

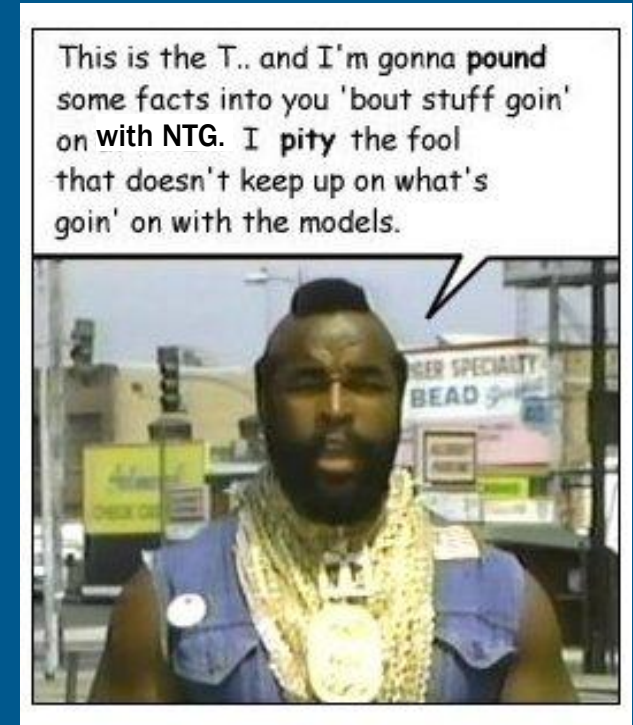
Evaluation Consultants



Components of NTG

- Net Savings adjust for:

- Free riders
- Participant Spillover
- Market effects



Net = (Gross – Freeriders + Participant Spillover + Market Effects)

Caveat: It has to be measured / documented in an independent program evaluation

Gross vs. Net Savings

Why all the fuss about gross vs. net

- Gross savings are easier to measure but may not reflect the actual savings induced by the program - policy makers not sure of actual impacts
- Net savings are difficult (and sometimes) impossible to accurately measure but in theory they provide a better measure of what effect the program had (as opposed to what would have been naturally occurring)

NTG Approaches

- **How do you measure program attribution?**
 - Deemed, or stipulated, net-to-gross ratios (least expensive/least precise)
 - Self-reporting surveys (moderately expensive)
 - Enhanced self-reporting surveys (moderately expensive, but more than self-report alone)
 - Econometric methods (more expensive)
 - Market based analysis (data intensive, market cooperation needed)
- **Best practices continually evolve, but typically include multiple methods or data sources**

– Are net assessment approaches reliable?

- Deemed, or stipulated, net-to-gross ratios
 - least expensive – can be adjusted over time and based on best information available
- Self-reporting surveys
 - Self selection / false response / positive outcome / bias
- Enhanced self-reporting surveys
 - does not solve the bias issue, just brings more of it in
- Econometric methods
 - Very difficult and expensive to do well
- Market based analysis
 - Often not granular enough and data can be difficult to get.

– Why climb this slippery slope? Additionality!

A process evaluation...

- **Systematically** reviews a single program or portfolio of programs
- **Regularly** provides feedback on program progress and performance
- **Openly** recognizes what is working well with a program and identifies program design issues and barriers to delivery
- **Clearly** lays out actionable recommendations for program improvements or changes in program goals

•And let's not forget, it often fulfills a **regulatory** requirement and is often directly or indirectly tied to recovery

What is Assessed

- Program design and processes
- Program administrative activities
- Program delivery and implementation activities
- Customer response
- Internal and external program barriers
- Market response
- Program impacts

Elements of a Good Process Evaluation

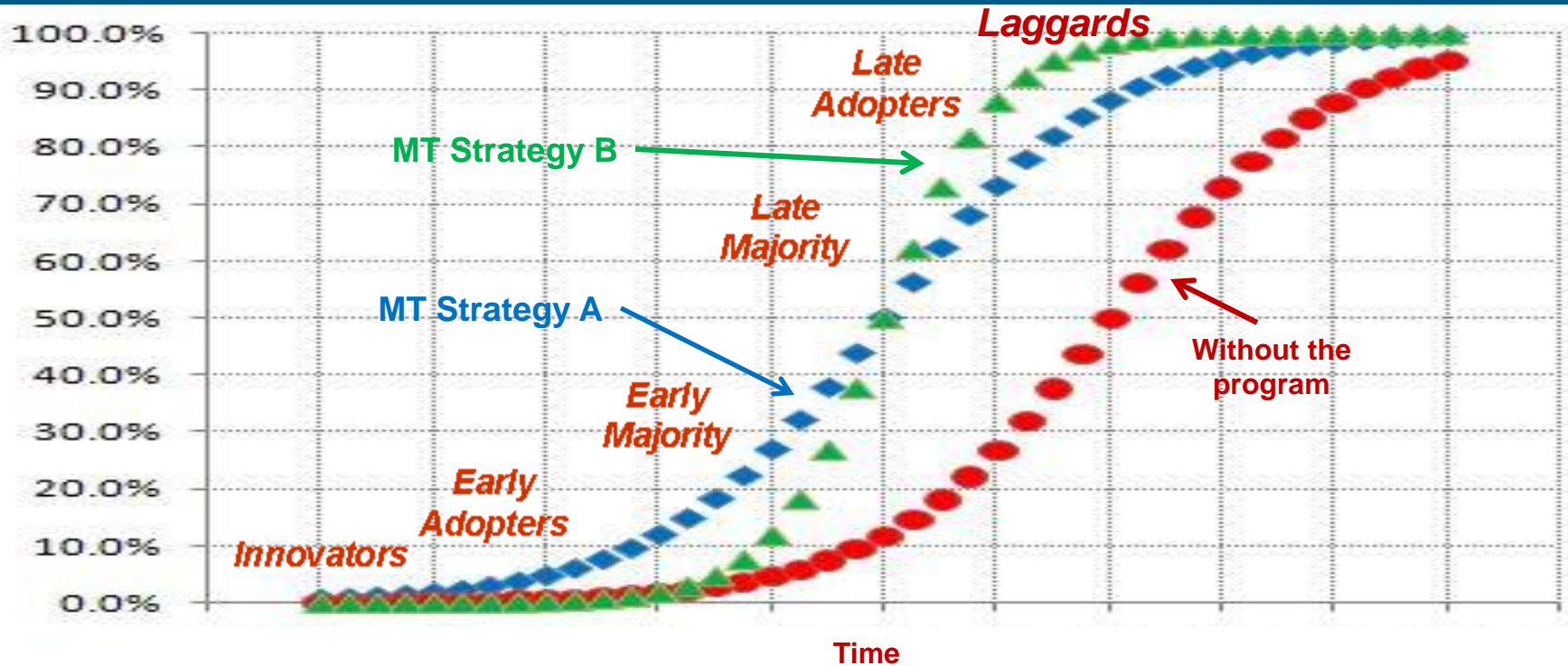
A good process evaluation...

- Is objective and unbiased
- Is systematic and timely
- Identifies what is working well and opportunities for improvement
- Maintains regular communication with stakeholders
 - Feedback should be provided after each major activity
 - Regular status updates should be communicated
 - Provide the option of a presentation to stakeholders
 - Share the findings with program managers and discuss implications and improvements in advance of next program planning cycles

Market Effects Evaluation

- Market transformation is present when a program design are effective at overcoming barriers to adoption within the marketplace.
- Market transformation can have big effects.
- Savings can be much larger than standard programs when successful (using the market expands reach sometimes without expanding program administrative costs)

ME Goal – early movement along the S-curve



Cost Effectiveness

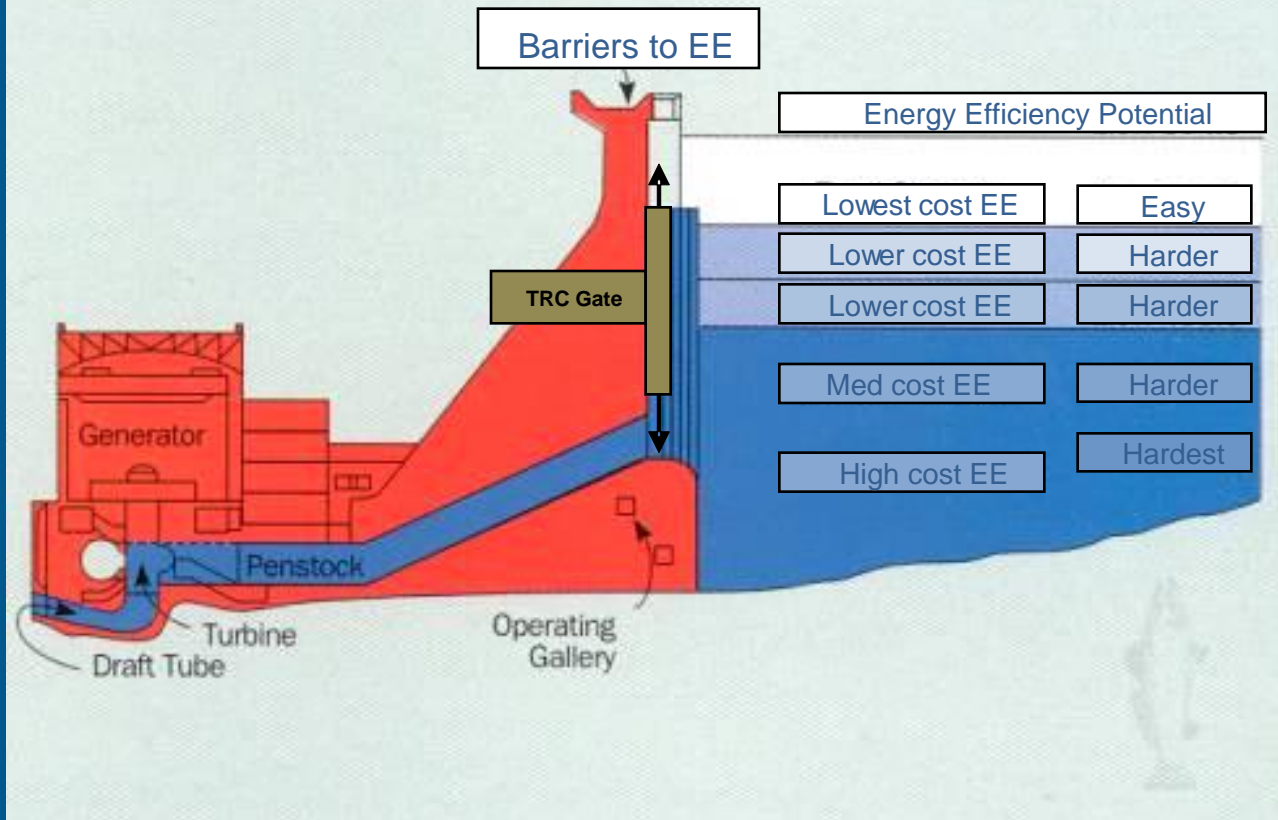
C/E Policy decisions drive everything

The TRC is the EE energy supply valve.

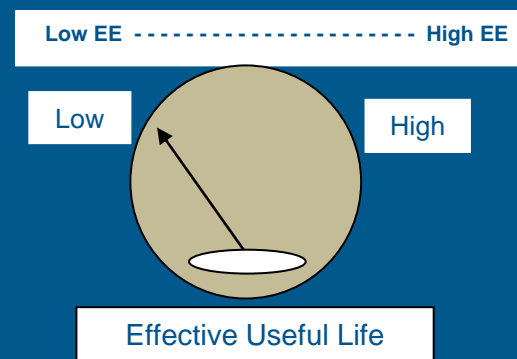
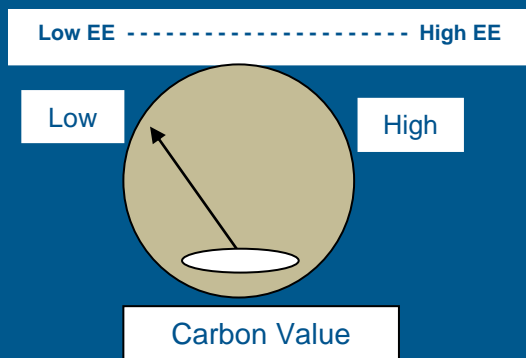
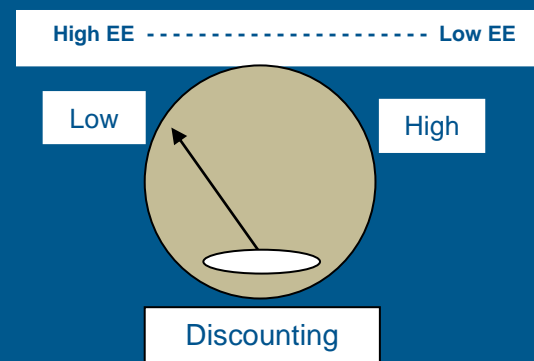
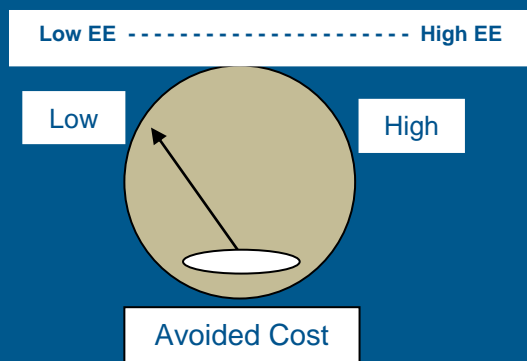
**How cost effectiveness test are set up
sets the limits of what can be achieved.**

The TRC is the EE Supply Valve

TRC opens and closes the EE accomplishment gate



The Control Valves are...



These values are set is policy decisions.
How they are set limits the amount of energy and carbon that can be saved.

Cost Effectiveness Tests

Test	Acronym	Key questions answered	Summary approach
Participant cost test	PCT	Will customers choose to participate?	Comparison of costs and benefits of the customer installing the measure. Ignores impact on utility, non-participants and society of making or not making the investment
Program administrator cost test / (Utility cost test)	PACT	Will utility bills in aggregate be lower?	Comparison of program administrator costs to supply-side resource costs
Ratepayer impact measure	RIM	Will energy efficiency contribute to utility rates increase?	Comparison of administrator costs and utility bill reductions to supply-side resource costs
Total resource cost test	TRC	Will the total costs of energy decrease?	Comparison of program administrator and customer efficiency costs to utility resource savings
Societal cost test	SCT	Is society better off as a whole?	Comparison of society's costs of energy efficiency to resource savings in addition to other societal costs and benefits

Different tests do different things

Cost Benefit Elements	TRC	RIM	UCT	PCT	SCT
Benefits					
Avoided Power Supply Costs	X	X	X		X
Avoided T&D Costs	X	X	X		X
Bill Reductions				X	
Non Energy Benefits					X
Costs					
Direct Utility DSM Costs	X	X	X		X
Direct Customer DSM Costs	X			X	X
Utility Program Administration	X	X	X		X
Lost Revenues		X			

Introducing Evaluation Plans



It is the Evaluation Road Map

- Provide overview of program by
 - Confirming we have an accurate understanding of program and its goals
- Documents research objectives, program metrics, and researchable issues
- Lays out research methodology, assumptions, and activities
- Ensures pressing research questions are included in activity
- Establishes a timeline and key deliverables

Evaluation Plan – Possible Contents

- Evaluation Goals
- Program Description
- Program Logic Model
- Key Researchable Questions
- Metrics to Be Measured
- Sampling/Data Collection Plan
- Cross-Cutting Evaluation Activities
- Impact Evaluation Approach
- Process Evaluation Approach
- Market Effects Assessment
- Budget and Schedule
- Report Content and Structure.



Evaluation Costs

- Generally, total evaluation costs range from three percent to six percent of program costs
- Most states set @ 5% for evaluation
- Budget drives reliability
 - What reliability do you want – need?
- Evaluation costs can vary depending upon the following elements:
 - Size of program
 - Maturity of program & past evaluation efforts
 - Purpose of the evaluation
 - Type of evaluation
 - Audience
 - Required level of statistical precision
 - Timing – how often you evaluate each program

Cost-Quality-Time Relationship



Quality



Time



Cost

Choose two!

Reporting Results

Standard format includes following pieces:

- Executive summary
- Introduction
 - Includes program description, researchable issues, and methodology.
- Key findings
- Conclusions and recommendations
- Appendices with documentation on sample design, response rate, data collection instruments, technical data, etc.

Wrap-up and Questions

